VISUAL SENSITIVITY ANALYSIS

(APPLIED TO REAL ESTATE PREDICATION SYSTEM)

A thesis submitted in fulfilment of the requirements for the degree of

Master of Science in Computing Sciences in the

Faculty of Engineering and information technology at

University of Technology, Sydney

MASSARA DA'ANA

JANUARY 2014

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

Acknowledgments

Many thanks to my supervisor, A/Prof. Mao Lin Huang, who encouraged this research idea from the very beginning. It would not have been possible to accomplish this work without his constant guidance, support and invaluable advice. Thank you again for teaching me how to become a better researcher.

Thanks to Mr. Mohammed Al Alkadi and Miss Zhenya Zhang for their help while developing the prototype system.

Thanks for Dr. Esam Dana and Dr. Zenat Dana for assisting me in different ways throughout this research.

Thanks to my husband, my sisters and brothers, for their constant encouragement and support, and to my children for respecting the time I spent on this research more than I expected. To them all, I dedicate this thesis.

Table of Contents

ACKNOWLEDGMENTS III
TABLE OF CONTENTS IV
LIST OF FIGURES VI
LIST OF TABLES VII
ABSTARCT VIII

1 INTRODUCTION 1

1.1 MOTIVATION 1
1.2 CONTRIBUTION 3
1.3 ORGANIZATION 4

2 RELATED WORK 5

2.1 SENSITIVITY ANALYSIS 5
2.1.1 SA Classification based on SA capabilities 6
2.1.2 SA Classification based on how the method is conducted 8
2.2 VISUALIZATION APPROACHES 23
2.2.1 Overview + Detail Approach 24
2.2.2 Focus + Context Approach 31
2.3 SUMMARY 36

3 RESEARCH METHODOLOGY AND SYSTEM DESCRIPTION 37

3.1 SCOPE 37
3.2 PART 1: PREDICTION SYSTEM DEVELOPMENT. 38
3.3 PART 2: MODEL DEVELOPMENT 46

4 VISUAL SENSITIVITY DATA ANALYSIS. 53

4.1 MODEL EXPLAINED VARIABILITY. 53
4.2 OVERALL MODEL SIGNIFICANCE. 53
4.3 INDIVIDUAL COEFFICIENT SIGNIFICANCE 54
4.4 IMPLEMENTATION WITH REAL ESTATE DATA. 56
4.4.1 Adjusting variable One with resulting visualization. 56
4.4.2 Adjusting variable Two with resulting visualization. 58
4.4.3 Adjusting variable three with resulting visualization. 60

5 CASE STUDIES 62

5.1 CASE STUDY 1: PREDICTION WHEN ADJUSTING NUMBER OF BED ROOMS. 62
5.1.1 Discussion 65
5.2 CASE STUDY 2: PREDICTION WHEN ADJUSTING DISTANCE TO SYDNEY CBD. 66
5.2.1 Discussion 67
5.3 CASE STUDY 3: PREDICTION WHEN ADJUSTING DISTANCE TO NEAREST TRANSPORT. 68
5.3.1 Discussion 69

6 COMPARATIVE STUDY 70

6.1 SYSTEM 1: USING LME ALGORITHM 70
6.2 SYSTEM 2: USING GAMMA TEST. 71
6.3 SYSTEM 3: DOMAIN.COM 73

7 CONCLUSION AND FUTURE WORK 76

7.1 CONCLUSION 76
7.1.1 Interactive Visualization *76*
7.1.2 Sensitivity Analysis *76*
7.2 FUTURE WORK. 76
APPENDIX 77
BIBLIOGRAPHY 79
PUBLICATIONS 81

List of Figures

FIGURE 1: TORNADO DIAGRAM. 12
FIGURE 2: RADAR CHART. 14
FIGURE 3: SCATTER PLOT OF PROPERTY PRICES (MILLION DOLLARS) VERSUS THE DISTANCE TO SYDNEY CBD (KM). 16
FIGURE 4 SCATTER PLOT MATRIX. 18
FIGURE 5: A PARALLEL COORDINATE PLOTS 20
FIGURE 6: GOOGLE MAPS 26
FIGURE 7: MICROSOFT POWERPOINT'S OVERVIEW AND DETAIL INTERFACE 27
FIGURE 8: A TREE DIAGRAM 28
FIGURE 9: A TREE-MAP FOR THE TREE IN FIGURE 8 28
FIGURE 10: MAP OF NEW SOUTH WALES 29
FIGURE 11: MAP OF SYDNEY WITH REMOTE SUBURBS 30
FIGURE 12: MAP OF THE CITY OF SYDNEY 30
FIGURE 13: TWO TYPICAL TECHNIQUES OF FOCUS + CONTEXT APPROACH 32
FIGURE 14: A MULTIPLE-FOCUS VIEW OF THE PROJECTION WITH THE SAME PARAMETERS FOR EACH FOCUS POINT (A) AND WITH VARIOUS PARAMETERS (B) 33
FIGURE 15: PROPOSED PROPERTY PREDICATION SYSTEM. 38
FIGURE 16: THE OVERVIEW OF IMAP 40
FIGURE 17: VISUAL QUERY OUTCOME, DETAILED PROPERTY FEATURES. 41
FIGURE 18: IMAP REVEALING FILTERED INFORMATION IN THE DETAILED AREA. THE OVERVIEW AREA IS ZOOMED OUT TO THE BOTTOM LEFT. 42
FIGURE 19 : IMAP REVEALING FILTERED INFORMATION IN THE DETAILED AREA. THE OVERVIEW AREA IS ZOOMED OUT TO THE BOTTOM LEFT. 43
FIGURE 20: COMPILED RESULTS VISUALIZATION. THE PREDICTED PRICE IS FOR THE PROPERTY WITH ONE UNIT INCREASE IN THE PREDICTOR "NUMBER OF BED ROOMS". 45
FIGURE 21 : A) SCATTER PLOT FOR THE RESPONSE VARIABLE AND CBD 48
FIGURE 22 : A) SCATTER PLOT MATRIX FOR THE TRANSFORMED RESPONSE AND ALL OTHER PREDICTORS. B) SCATTER PLOT MATRIX FOR THE RESPONSE IN ITS ORIGINAL FORM AND ALL OTHER PREDICTORS. 50
FIGURE 23: COMPILED RESULTS, NUMBER OF BED ROOMS VS. PREDICTED AND OBSERVED PRICE 57
FIGURE 24: COMPILED RESULTS, DISTANCE TO SYDNEY CBD VS. PREDICTED AND OBSERVED PRICE . 59
FIGURE 25: COMPILED RESULTS, DISTANCE TO NEAREST TRANSPORT (DTRAIN) VS. PREDICTED AND OBSERVED PRICE 61
FIGURE 26: THE SELECTED REGION "FAIRFIELD AND LIVERPOOL", WITH THE PROPERTY OF INTEREST HIGHLIGHTED. THE TOOLTIP SHOWS SOME OF THE PROPERTY FEATURES. 63
FIGURE 27: THE PROPERTY FEATURES FORM SHOWS ALL AVAILABLE INFORMATION ABOUT THE PROPERTY OF INTEREST FOR THE CANAANS. 64
FIGURE 28: "CUSTOMIZE PROPERTY" FORM. THE DROP DOWN LIST SHOWS THE FEATURES THAT THE USER CAN ADJUST PRIOR TO THE ESTIMATION OF THE PRICE. 65
FIGURE 29: THE PREDICTED PRICE FOR THE PROPERTY WITH ADJUSTED BEDR. 65
FIGURE 30: A TOOLTIP SHOWS THE SUMMARY OF PROPERTY FEATURES 66
FIGURE 31: THE PREDICTED PRICE AFTER ADJUSTING CBD PREDICTOR. 68
FIGURE 32: THE PREDICTED PRICE AFTER ADJUSTING DTRAIN PREDICTOR. 69

List of Tables

TABLE 1: COMPARISON OF VISUALIZATION TECHNIQUES 35
TABLE 2: REVIEWED ALGORITHMS/SYSTEMS FOR PROPERTY PRICE PREDICTION FEATURES SUMMARY. 75

ABSTRACT

Sensitivity analysis is the science studies the impact of independent variables on the dependant variable in the studied model, in addition to investigating relationships between those variables. Sensitivity analysis is a prevalent group of techniques and approaches has proven its feasibility in wide range of disciplines.

However, the traditional sensitivity analysis methods have the common weakness of *user interaction absence*. Furthermore; each sensitivity analysis method has its *own level of difficulty which is an obstacle for a non-expert user* to use or even to interpret the results if the analysis is conducted using a software like SPSS.

Recently, visualizations are being used to present data efficiently in terms of assisting human visual perception and reducing cognition effort. These visualization techniques when integrated with interaction will facilitate data manipulation and exploration.

This study integrates sensitivity analysis with interactive visualization into a prediction system that allows non-expert users to analyse and understand the real estate data through the visualization and direct visual interactions, which hide the complexity of the sensitivity analysis algorithms. We take advantage of the visualization that amplifies cognition in dealing with abstract data.

As shown in the outcome, the user can use the sensitivity analysis method used in this system *interactively* setting his/her preferences for the property via the visualization *without any prerequisite of sensitivity analysis knowledge.*

The use of scatter plots, one of sensitivity analysis methods, is used in studying the relationships between the predictors and the response variables to decide whether variable transformation is needed or not. Additionally, scatter plots are used to summarize all analyses conducted.